PhytoKeys

PhytoKeys 249: 95–114 (2024)
DOI: 10.3897/phytokeys.249.133205

Research Article

Morphological and molecular characterization of *Halamphora vantushpaensis* (Bacillariophyceae, Amphipleuraceae), a new diatom species widely dispersed on the shores of the soda Lake Van (Türkiye)

Elif Yılmaz[1], Romain Gastineau[1], Cüneyt Nadir Solak[2], Ewa Górecka[1], Rosa Trobajo[3], Monique Turmel[4], Claude Lemieux[4], Christian Otis[5], Andrzej Witkowski[1†], David G. Mann[3,6]

1 *Institute of Marine and Environmental Sciences, University of Szczecin, Mickiewicza 16A, PL70–383 Poland*
2 *Department of Biology, Faculty of Science & Art, Dumlupınar University, 43000 Kütahya, Turkiye*
3 *Marine and Continental Waters, Institute for Food and Agricultural Research and Technology (IRTA), Crta de Poble Nou Km 5.5, E-43540 La Ràpita, Catalunya, Spain*
4 *Département de biochimie, de microbiologie et de bio-Informatique, Institut de Biologie Intégrativeet des Systèmes, Université Laval, Québec, QC, Canada*
5 *Plateforme d'Analyse Génomique, Institut de biologie intégrative et des systèmes, Université Laval, Québec, QC, Canada*
6 *Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, Scotland, UK*

Corresponding authors: Elif Yılmaz (elfyilmaz38@gmail.com); Romain Gastineau (romain.gastineau@usz.edu.pl)

Abstract

In this study, we describe *Halamphora vantushpaensis* **sp. nov.**, a newly identified diatom species found in the highly alkaline Lake Van in Eastern Turkey (Türkiye). This new species is characterized morphologically by light and scanning electron microscopy, performed on both wild and cultivated samples. Two monoclonal cultures were submitted to a genome-skimming approach, giving access to the complete sequence of their nuclear rRNA cluster of genes, mitochondrial and plastid genomes. Both strains were highly similar from the genomic point of view, with few mutations noted, although in organellar genomes some of them concerned protein coding genes and were non-silent. Also, the group II intron in the mitochondrial *cox1* gene was found to display a relatively high number of polymorphisms. The plastome also distinguishes itself from other *Halamphora* spp. by the extension of its inverted repeat at the expense of the two single copy regions of the genome. Maximum likelihood molecular phylogeny inferred from a concatenated three genes dataset *(18S*, *psbC* and *rbcL)* positions this species within the K clade, which is known to contain hypersaline to freshwater species.

Key words: Alkaline lake, group II introns, inverted repeat, mitochondrial genome, multigene phylogeny, nuclear rRNA genes, plastid genome

OPEN ACCESS

Academic editor: Kalina Manoylov
Received: 26 July 2024
Accepted: 4 September 2024
Published: 15 November2024

Citation: Yilmaz E, Gastineau R, Solak CN, Górecka E, Trobajo R, Turmel M, Lemieux C, Otis C, Witkowski A, Mann DG (2024) Morphological and molecular characterization of *Halamphora vantushpaensis* (Bacillariophyceae, Amphipleuraceae), a new diatom species widely dispersed on the shores of the soda Lake Van (Türkiye). PhytoKeys 249: 95–114. https://doi.org/10.3897/phytokeys.249.133205

Copyright: © Elif Yılmaz et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Soda lakes are among the rarest and most geochemically distinctive wetlands on Earth. They are characterized by their alkaline waters containing high levels of carbonate and bicarbonate ions, typically resulting in elevated pH levels. Lake Van is the largest soda lake in the world (Glombitza et al.

† Deceased author.

2013), with water that is both saline (21.4‰) and alkaline (155 m mEq^{-1}, pH 9.81) (Aydar et al. 2003; Kempe et al. 1991). The lake has existed for 600,000 years, spanning multiple glacial–interglacial cycles (Stockhecke et al. 2014; North et al. 2018) and hosts endemic species of fishes (e.g. Akku et al. 2021). However, studies on the phytoplanktonic flora of the lake, and especially diatoms, have been rather scarce.

For a long time, investigations into the contemporary diatom flora from Lake Van were restricted to a single study by Legler and Krasske (1940). These authors worked on samples brought to them in Germany from Lake Van and, based on light microscopy, they found 24 diatom taxa. Among them, two belonged to the genus *Amphora*. The first one was described as '*Amphora coffeiformis* Ag.' (more correctly referred to as *Amphora coffeiformis* (C.Agardh) Kützing, 1844). This is now a non-accepted synonym of *Halamphora coffeiformis* (C.Agardh) Levkov following the recent revisions of the amphoroid diatoms in which Cleve's sect. *Halamphora* has been recognized at the genus level (Cleve 1895; Levkov 2009; Stepanek and Kociolek 2013, 2014, 2019). The second was '*Amphora commutata* Grun.', published in Van Heurck (1880), considered a valid species as of today. Both taxa are known for being brackish species and both have wide distributions (Guiry and Guiry 2024).

In the last few years, new investigations have been conducted on Lake Van diatoms using an integrative approach that combines light/scanning electron microscopy (LM/SEM) and molecular phylogenies derived from next generation sequencing results. With these data, three new species already have been discovered and described, namely *Nitzschia anatoliensis* Górecka, Gastineau and Solak (Solak et al. 2021), *Navicula vanseea* Yılmaz, Gastineau, Solak and Witkowski (Yilmaz et al. 2024) and *Halamphora witkowskii* (Yilmaz et al. in press). Lake Van can be divided into four different basins of various depths: the shallow Erçis basin (northeast), the Van basin (southeast), the Ahlat basin (northwest) and, at the centre of the lake, a deeper fourth basin (Kaden et al. 2010). *Navicula vanseea* and *Ni. anatoliensis* were both described from material from the Erciş basin but have not been observed yet in other parts of the lake. *Halamphora witkowskii* is known so far only from the Ahlat basin (Yilmaz et al. in press).

In the current article, we describe another *Halamphora* species, *Halamphora vantushpaensis* sp. nov., using the same integrative approach and tools previously used for *Na. vanseea* and *Ni. anatoliensis*.

Material and methods

Sampling, isolation and cultivation

Live samples were collected by scraping stones at four different stations around Lake Van: Ahlat, Bitlis (38°75'45.748"N, 42°50'71.257"E); Erciş, Van (39°00'07.9"N, 43°25'40.4"E); Adilcevaz, Bitlis (38°79'83"N, 42°72'16"E); and Edremit, Van (38°42'07.09"N, 43°23'74.39"E) (Fig. 1). Individual diatom cells were isolated from the Ahlat samples using a micropipette under a Nikon TS100 inverted microscope (NIKON, Tokyo, Japan). The strains were subsequently moved into 250 mL Erlenmeyer flasks containing F/2 medium (Guillard, 1975), which had been adjusted to 18‰ salinity. The cultures were maintained in conditions promoting active growth, with a light intensity of 60 µmol photons m^{-2} s^{-1},

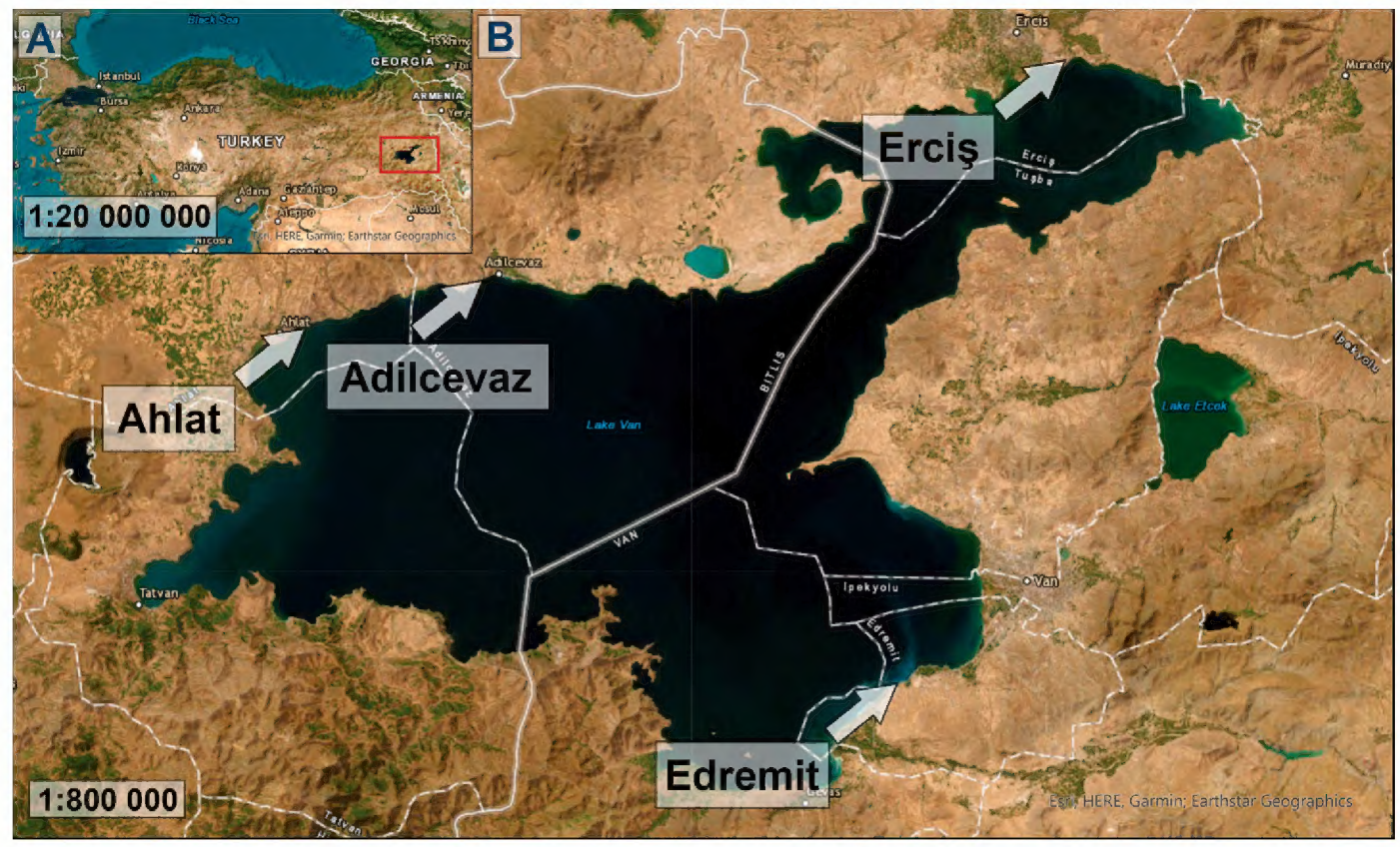

Figure 1. Map of the sampling location **A** location of Lake Van in Turkey. The red frame indicates the position of Lake Van **B** general view of the lake. The areas indicated by arrows are the sampling stations. (Esri. (2023). ArcGIS Pro 3.1.0. Environmental Systems Research Institute)

and a photoperiod of 14 hours of light and 10 hours of darkness at a temperature of 18 °C. The two monoclonal cultures were obtained and registered in the Szczecin Diatom Culture Collection as SZCZEY 2166 and SZCZEY 2167.

Light and scanning electron microscopy

Light microscopy (LM) documentation was obtained at the University of Szczecin with a Zeiss Axio Scope A1 (Carl Zeiss, Jena, Germany) using a Canon EOS 500D camera and Canon EOS Utility software (Canon, Tokyo, Japan). Images were obtained using a 100× Plan Apochromat oil immersion objective (numerical aperture = 1.4).

For the preparation of diatom frustules for both LM and scanning electron microscope (SEM) observations, samples (pellets of cells from monoclonal cultures or wild samples) were moved into 20 mL beakers and 10 mL of 10% HCl added. Over a 24-hour period, the samples were washed with distilled water four times, allowing the material to sediment naturally between washes. Next, the samples were re-suspended in 30% H_2O_2 and boiled on a hotplate at 170 °C for approximately four hours. The final step involved washing the samples four times with distilled water in 24 h, as before. For LM, the material was air-dried on cover glasses and then affixed to a glass slide using Naphrax (Brunel Microscopes, Chippenham, UK). LM measurements were done on a total of 91 valves. For SEM, a drop of cleaned sample was placed on a Nuclepore Track-Etch membrane (Whatman, Maidstone, U.K.). Following air-drying overnight, the membranes were mounted on aluminium stubs using carbon tape and coated with gold using a Q150T coater (Quorum Technologies, Laughton, UK). SEM imaging was conducted at the Faculty of Chemical Technology and Engineering, Western Pomeranian University of Technology in Szczecin (Poland) on a Hitachi SU8020 field emission microscope (Tokyo, Japan). The imaging was conducted with an accelerating voltage of 5kV and a working distance of 8500–8600 µm.

Next generation sequencing and bioinformatic analyses

Clones SZCZEY2166 and SZCZEY2167 were harvested by centrifugation and DNA was extracted following Doyle and Doyle (1990). Total DNA was sent to the Beijing genomics Institute (BGI) in Shenzhen (China) where they were sequenced on a DNBSEQ platform for a total for each clone of ca. 40M 150 bp paired-end clean reads. Assemblies were conducted using SPAdes 3.15.0 (Bankevich et al. 2012) with a k-mer parameter of 125. The contigs corresponding to the plastid and mitochondrial genomes or nuclear rRNA clusters were data-mined by standalone blastn queries. The subunits of the plastid genome were merged with each other with the help of Consed (Gordon and Green 2013). Annotation of protein coding genes was done as described in Gastineau et al. (2023).

Molecular phylogeny

Maximum likelihood phylogenies were inferred from two different datasets. One contained a concatenated alignment of *18S*, *rbcL* and *psbC* genes representing 214 taxa downloaded from GenBank appended with those obtained in the course of this study. Two strains of *Triparma pacifica* (Guillou and Chrétiennot-Dinet) Ichinomiya and Lopes Santos were used as an outgroup. Among the diatom taxa, two were lacking *rbcL* data, 22 strains lacked *psbC* and 21 strains lacked 18S. A *rbcL*-only tree was built as well, in order to compare the tree topologies. Sequences were aligned using MAFFT 7 (Katoh and Standley 2013) and trimmed automatically with trimAl (Capella-Gutiérrez et al. 2009). The best model of evolution was selected separately for each gene with ModelTest-NG (Darriba et al. 2020). In case of three-gene dataset, trimmed alignments of *18S*, *rbcL* and *psbC* were concatenated with Phyutility 2.7.1 (Smith and Dunn 2008). Maximum Likelihood phylogenes were computed using IQ-TREE 2.2.0 (Minh et al. 2020) with 1000 ultrafast bootstrap replicates; the dataset was partitioned based on the best models of evolution found for each gene. The trees were visualised with MEGA11 (Tamura et al. 2021). Lists of the sequences with their corresponding accession numbers can be accessed as explained in the data availability statement.

Results

Taxonomy

Phylum Bacillariophyta Haeckel
Class Bacillariophyceae Haeckel
Family Amphipleuraceae Grunow
Genus *Halamphora* (Cleve) Levkov

***Halamphora vantushpaensis* Yilmaz, Solak & Gastineau, sp. nov.**
Figs 2–4

LM (Figs 2A–M). Valves semi-lanceolate, dorsiventral with arched dorsal margin and slightly tumid ventral margin. Valve ends protracted and capitate in larger specimens (Figs 2A–F); but less protracted and not clearly separated

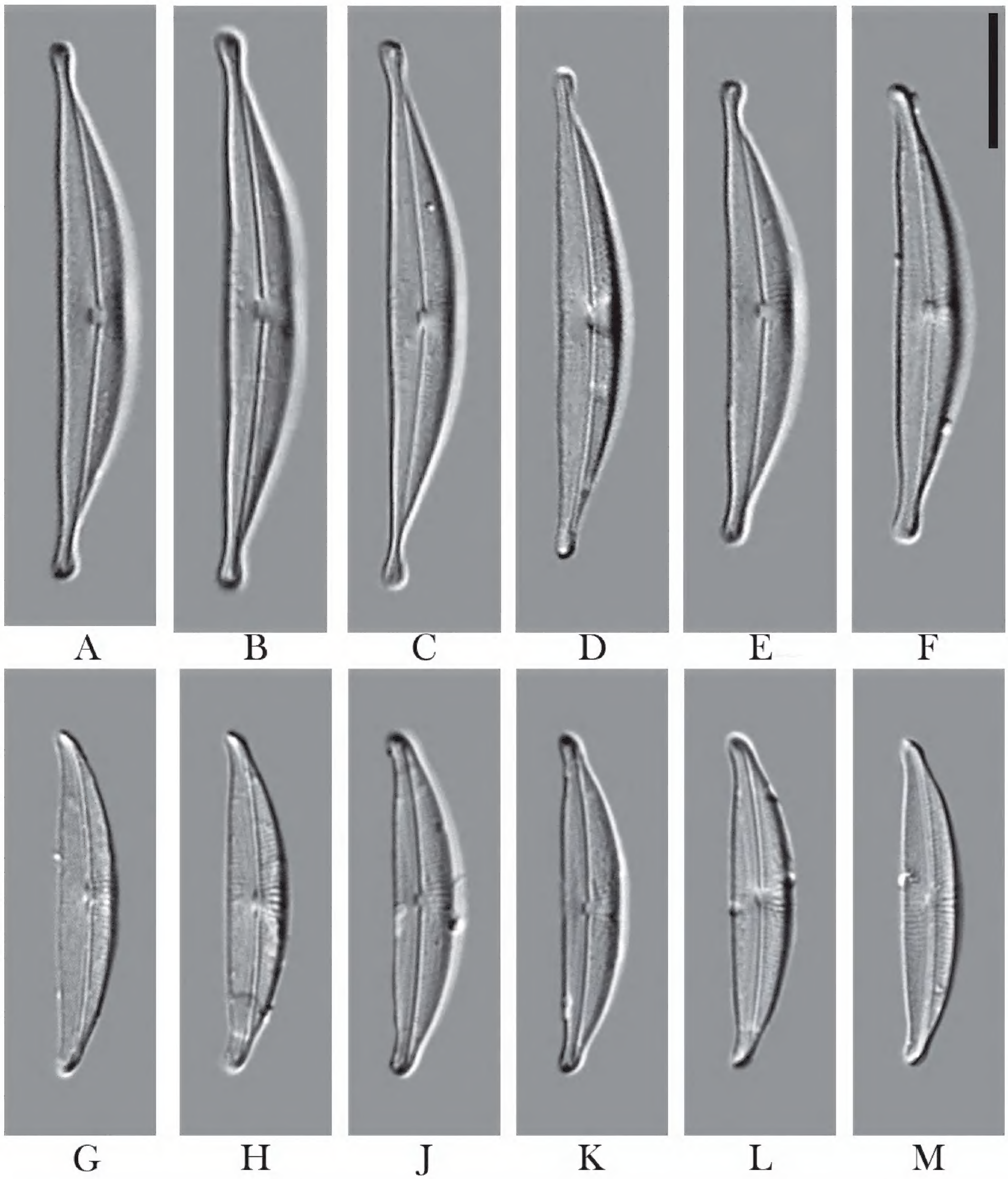

Figure 2. A–M *Halamphora vantushpaensis* sp. nov. LM micrographs **A–F** cleaned valves of the larger specimen (SZCZEY2167) **G–M** cleaned valves of the smaller specimen (SZCZEY2166). Scale bar: 10 µm

from the rest of the valve in smaller specimens (Figs 2G–M), ventrally bent. Valve length 24.0–42.0 µm, valve width 4.0–5.0 µm (n = 35). Axial area very narrow, wider on the ventral side. Central area visible in larger specimens: indistinct on the dorsal side, semi-lanceolate on the ventral side. Raphe almost straight, slightly arched, appearing to be located near the median line of the valve or slightly dorsal in valve view (Fig. 2). Sometimes the proximal raphe

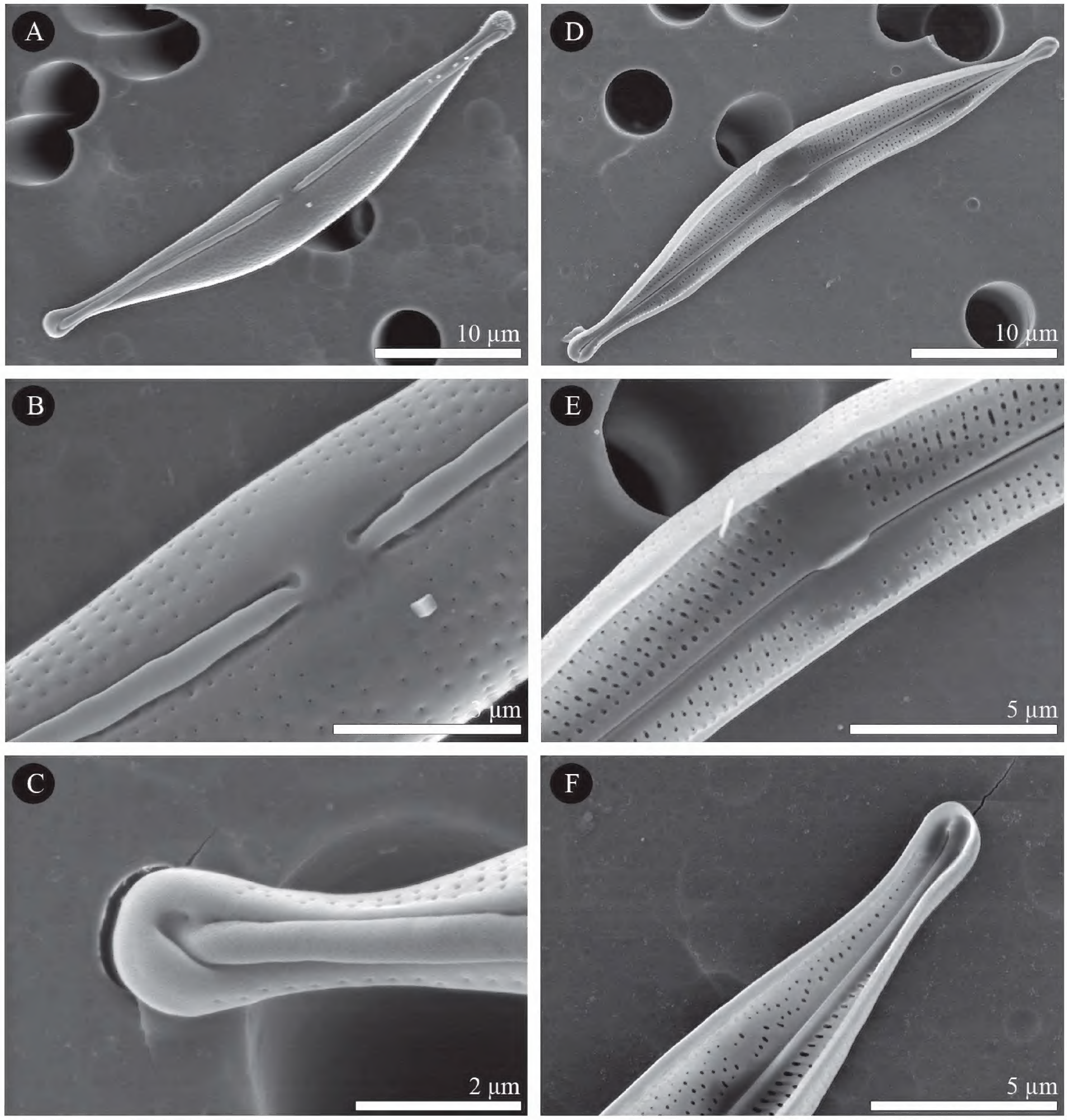

Figure 3. A–F *Halamphora vantushpaensis* sp. nov. SEM micrographs of strain SZCZEY2167 **A** External view of the entire valve **B** details of central area showing simple proximal raphe endings and regular shortened striae **C** details of apex showing the terminal fissure **D** internal view of the entire valve **E**details of central area showing fused central helictoglossae in proximal raphe endings **F** details of apex showing well-developed helictoglossae. Scale bars: 10 µm (**A, D**); 5 µm (**E, F**); 3 µm (**B**); 2 µm (**C**).

endings can be seen to be slightly dorsally bent (Fig. 2B). Striae hard to resolve in LM, dorsally slightly radiate over the entire valve (see SEM images for clearer demonstration of this), 27–32 in 10 µm.

SEM (Figs 3A–F, 4A–F, 5A–J). Externally, the valve face is arched, merging gently into the mantles (Figs 3B, 4B, C, 5H). Raphe ledge narrow and linear, present on the dorsal side of the raphe, with a prominent groove separating it from the valve face. The proximal raphe endings are slightly expanded into central depressions

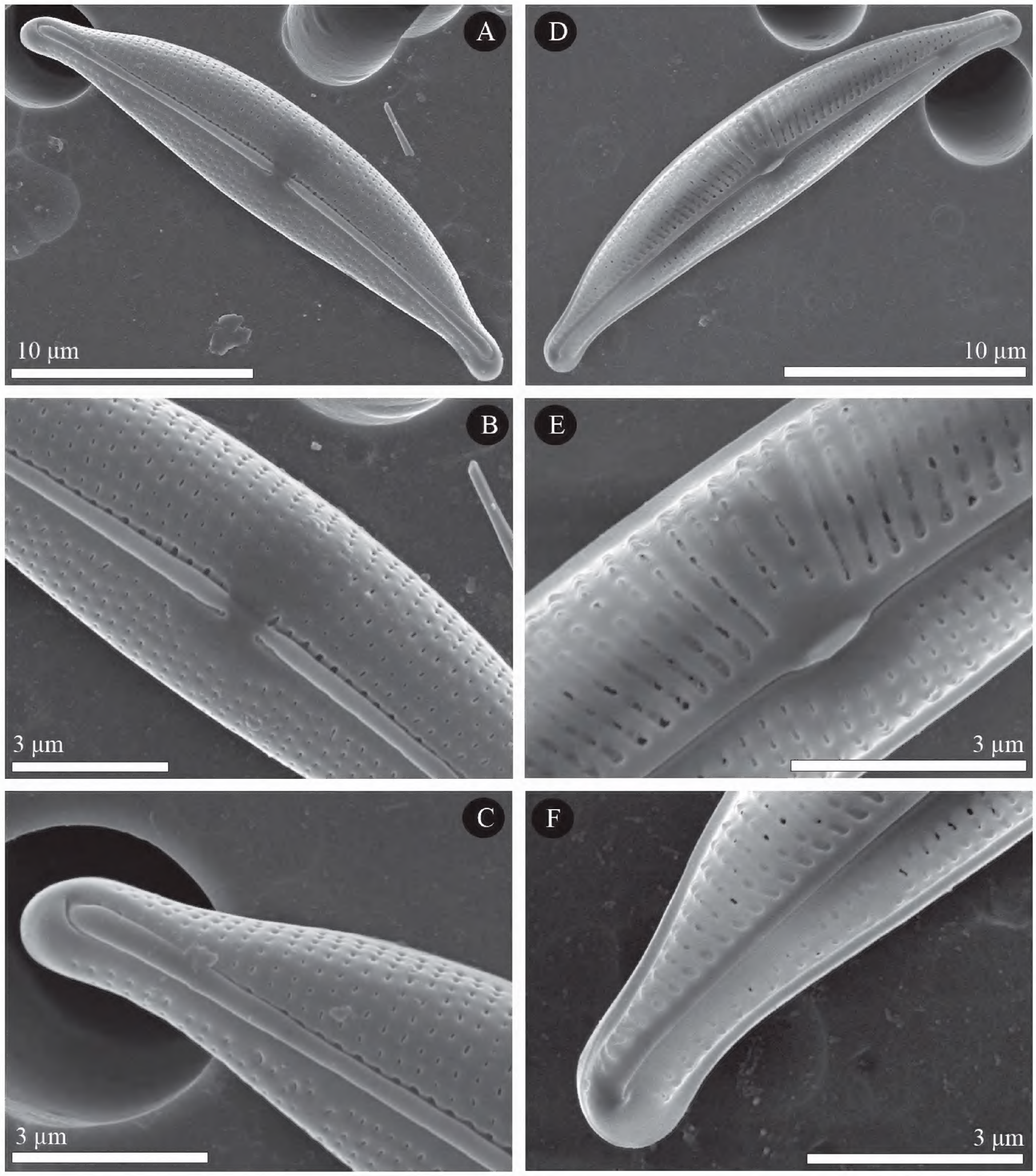

Figure 4. A–F *Halamphora vantushpaensis* sp. nov. SEM micrographs of strain SZCZEY2166 **A** external view of the entire valve **B** details of central area showing simple proximal raphe endings and regular shortened striae **C** details of apex showing the terminal fissure **D** internal view of the entire valve **E** details of central area showing fused central helictoglossae in proximal raphe endings **F** details of apex showing well-developed helictoglossae. Scale bars: 10 µm (**A, D**); 3 µm (**B, C, E, F**).

and are dorsally deflected (Figs 3B, 4B, 5H). The distal raphe endings are dorsally deflected and hook around to link with the groove bordering the raphe ledge (Figs 3C, 4C, 5J). The striae are simple and uniseriate, containing small round or slightly elongate poroids (Figs 3B, 4B, C, 5H), which are somewhat irregularly spaced (Fig. 4B, 5H and see also the internal views in Figs 3E, 4E, 5B, E).

The internal view of the valve shows the overall structure (Figs 3D, 5A). The central area is easier to detect than the external area and appears symmetrical

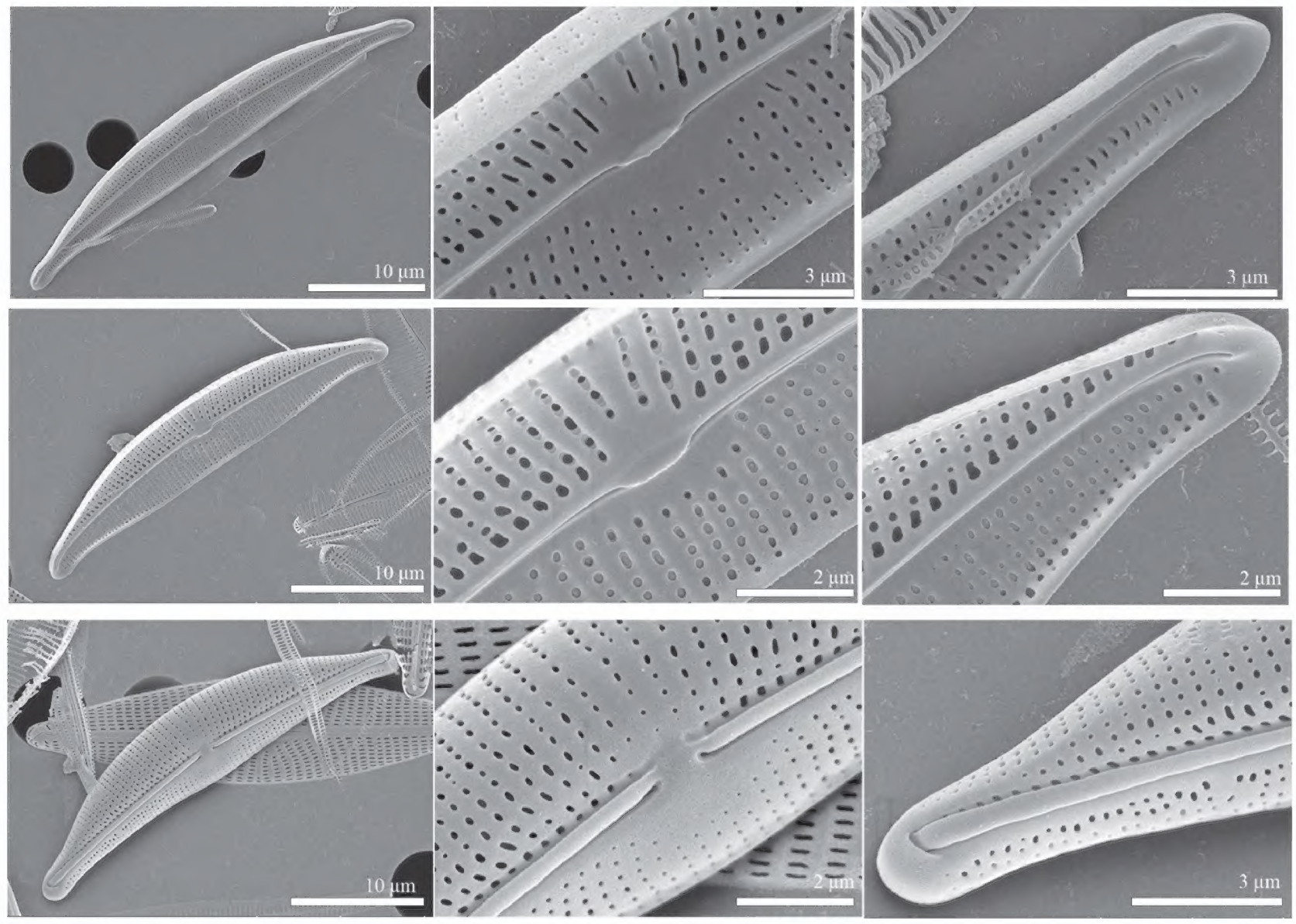

Figure 5. A–J SEM images of a cleaned valve from wild material **A** external view of the entire valve of the larger specimens **B** details of central area showing simple proximal raphe endings and regular shortened striae **C** details of apex showing the terminal fissure **D** internal view of the entire valve of the smaller specimens **E** details of central area showing fused central helictoglossae in proximal raphe endings **F** details of apex showing well-developed helictoglossae **G** external view of the entire valve of the smaller specimens **H** details of central area showing simple proximal raphe endings and regular shortened striae **J** details of apex showing the terminal fissure. Scale bars: 10 µm (**A, D, G**); 3 µm (**B, C, J**); 2 µm (**E, H, F**).

and large on the dorsal side in larger specimens (Fig. 3E, 5B); but very small on both sides in smaller specimens (Fig. 4E, 5E). Proximally, the raphe terminates within a fused central helictoglossa (Figs 3E, 4E, 5E). The distal raphe endings are slightly deflected ventrally and terminate in well-developed helictoglossae (Figs 3F, 4F, 5C, F). Internally, the poroids have round to elliptical internal openings (Figs 3E, 4E, 5B, E). These characteristics are summarized and compared with those of similar species in Table 1.

Phycobank. http://phycobank.org/104935.

Holotype. Slides number SZCZEY2167 in the collection of Andrzej Witkowski at the University of Szczecin, Poland. Valves representing the holotype population here illustrated in Fig. 2D.

Isotype. Slide number TR_Erciş_Van_2021 deposited in Kütahya Dumlupınar University (Türkiye).

Type locality. Erciş Van, Turkey (39°00'07.9"N, 43°25'40.4"E) leg. Elif Yılmaz, 31 July 2021.

Etymology. The species is named with regard to both Lake Van and the city of Tushpa, capital of the Iron Age kingdom of Urartu, which was located in the vicinity of the lake.

Distribution. The presence of this taxon has been assessed and confirmed at four different stations around Lake Van: Ahlat (North West of the lake), Adilcevaz (North), Erciş (North East), and Edremit (South East).

Table 1. Morphological characteristics of *Halamphora vantushpaensis* sp. nov. and similar *Halamphora* (-- represents no information) (for *H. vantushpaensis* measurements, n = 35).

	***Halamphora vantushpaensis* sp. nov.**	***H. atacamana***	***H. borealis***	***H. gasseae***	***H. salinicola***	***H. sardiniensis***	***H. thermalis***
Valve length (µm)	**24.0–42.0**	29.0–45.0	19.0–40.0	19.0–35.0	20.0–34.0	13.0–27.5	18.0–40.0
Valve width (µm)	**4.0–5.0**	4.5–8.0	3.0–4.0	3.5–4.5	2.5–3.7	3.0–4.5	4.0–6.5
Stria density (in 10 µm)	**27–32**	25–28	20–24	20–24	21–26	36–42	26–36
Valve outline	**semi-lanceolate with arched dorsal margin, slightly tumid ventral margin**	semi-lanceolate, arched dorsal margin, concave or straight to weakly tumid ventral margin	semi-lanceolate, arched dorsal margin, straight to weakly tumid ventral margin	semi-lanceolate, smoothly arched dorsal margin, straight to weakly concave ventral margin	semi-lanceolate, smoothly arched dorsal margin, straight to weakly concave ventral margin	semi-lanceolate, strongly arched dorsal margin and straight to slightly concave ventral margin	semi-lanceolate to lanceolate, smoothly arched dorsal margin, straight to slightly tumid ventral margin
Valve ending	**subprotracted in smaller specimens; protracted, capitate in larger specimens and ventrally bent**	slightly subprotracted and ventrally bent	protracted, capitate and slightly ventrally bent	shortly protracted and capitate	shortly protracted and capitate	shortly protracted and capitate	attenuate
Raphe ledge	**narrow, arched with equal width throughout**	narrow, equal width throughout and dorsally elevated from the valve face	narrow, linear	--	narrow, expanded on both valve sides	narrow, expanded on both valve sides	narrow, equal width throughout
Axial area	**narrow, widening ventrally**	narrow, widening ventrally	narrow, widening ventrally	narrow, widening ventrally	narrow, widening ventrally	narrow, widening ventrally	narrow, slightly dorsally bent
References	**in this study**	Levkov 2009	Levkov 2009	Levkov 2009	Levkov 2009	Levkov 2009	Levkov 2009

Genomics - the cluster of nuclear ribosomal RNA genes

Complete clusters of the rRNA genes, containing *18S*, internal transcribed spacer 1 (ITS1), *5.8S*, internal transcribed spacer 2 (ITS2) and *28S*, were obtained for both strains and registered with GenBank accession numbers PP726705 and PP726703 for SZCZEY2166 and SZCZEY2167 respectively. Their sizes and sequences were identical except for a single T/G SNP in the ITS1. The sizes of the different parts of the cluster are indicated in Table 2 and compared with results obtained on the same set of species as in Hamsher et al. (2019), which concern *Halamphora americana* Kociolek, 2014, *Halamphora calidilacuna* Stepanek & Kociolek, 2018 and *H. coffeiformis*. Lengths of *18S*, *5.8S* and *28S* were very conserved among species, except for *H. americana*, which has a group II intron in its *18S* that also contains on ORF coding for a putative reverse transcriptase protein. *Halamphora vantushpaensis* sp. nov. displays a longer ITS1 when compared to other species.

Genomics - mitochondrial genome

Complete mitogenomes were obtained on both strains and registered with GenBank accession numbers PP962256 (SZCZEY2166) (Fig. 6) and PP962257 (SZCZEY2167) (Fig. 7). The genomes are 42,659 bp and 43,152 bp long (SZCZEY2166 and SZCZEY2167 respectively). The genomes both contain 35 conserved protein coding genes (PCGs), two rRNA and 26 tRNA genes. The mitogenomes

Table 2. Length (in bp) of the different portions of the nuclear rRNA cluster for four species of *Halamphora* spp. The length of the *18S* gene of *H. americana* is indicated with and without the intron.

Species	*Halamphora vantushpaensis*	*Halamphora calidilacuna*	*Halamphora americana*	*Halamphora coffeiformis*
Accession number	PP726705 and PP726703	MH810165	MH810166	MH810167
Total length	5932	5764	9254	5938
18S	1767	1769	1783 (5241 with intron)	1767
ITS1	431	223	229	347
5.8S	156	155	155	154
ITS2	368	405	419	454
28S	3210	3212	3210	3217

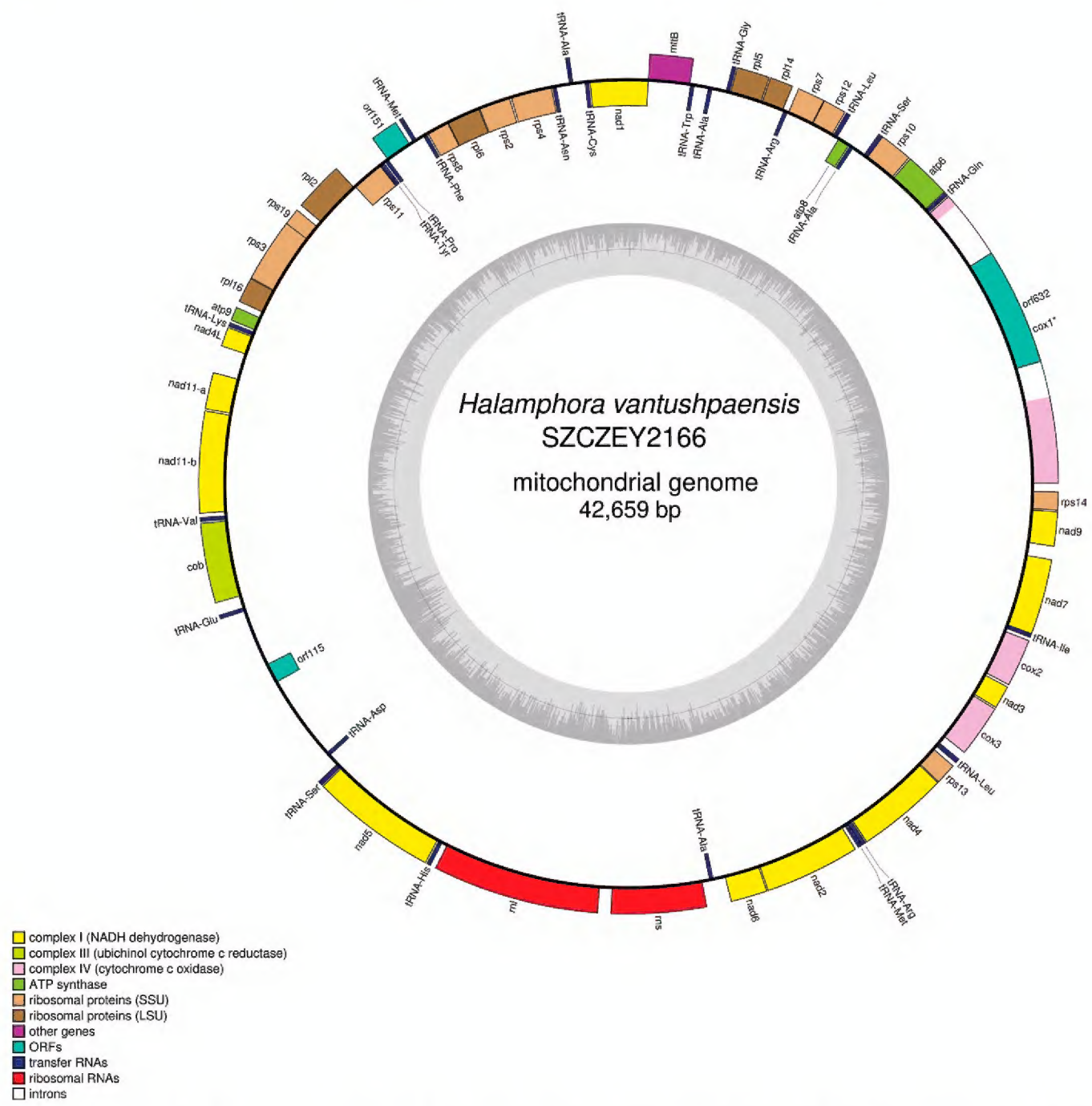

Figure 6. Map of the mitochondrial genome of *Halamphora vantushpaensis* sp. nov. SZCZEY2166

harbour the conserved open reading frame (ORF) orf151, although its position differs from most know species among which it is included in a conserved cluster of genes together with *rps11* and *mttb/tatC* (Pogoda et al. 2019) whereas here it is located between *rps11* and *rps8*. There is also a non-conserved ORF (orf115) between *cob* and *nad5*. The *cox1* gene contains a group II intron with an ORF coding for a putative reverse transcriptase. There are discrepancies in the length of this putative protein which is 632 amino-acid long in SZCZEY2166

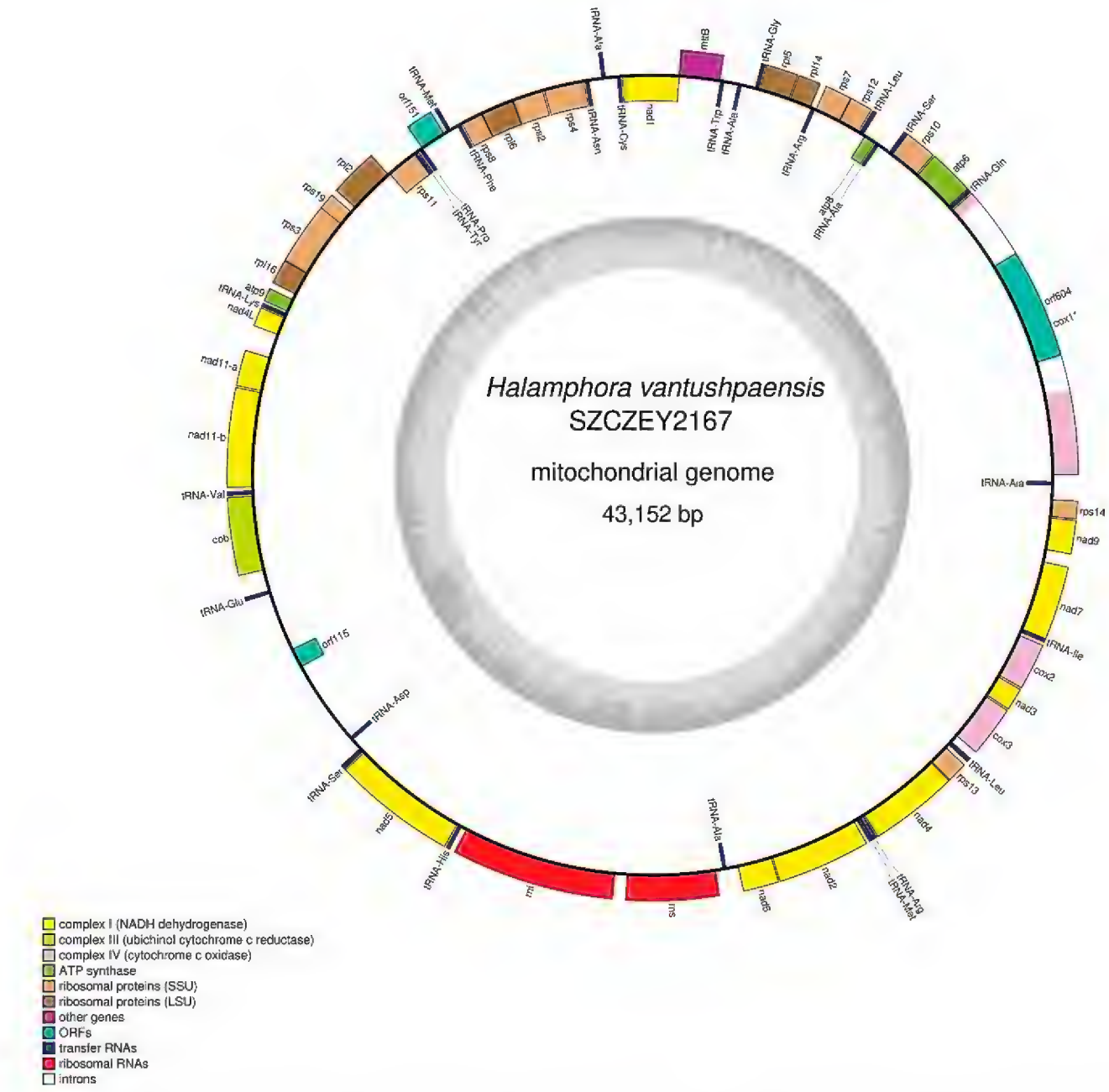

Figure 7. Map of the mitochondrial genome of *Halamphora vantushpaensis* sp. nov. SZCZEY2167

while it is 604 amino-acid long in SZCZEY2167, the extra-length being at the C-terminal part of the putative protein entirely. The polymorphisms between both strains mostly occurred in intergenic parts, hence the slight differences in lengths of the mitogenomes. The conserved protein coding genes were strongly conserved with some of them completely identical, although a certain number of polymorphisms could still be spotted in PCGs in the following genes, with the number of SNPs/lengths indicated between brackets: *cob* (6/1287), *nad2* (1/1536), *nad4* (1/1473), *rpl2* (1/810), *rpl6* (1/573), *rps3* (2/1038), *rps10* (1/540). These mutations were silent in *nad4* and *rpl2*, but led for each of the other encoded protein to one amino-acid substitution. Several variations were otherwise found in the *cox1* intron, whose size varied because of indels (four in total). It otherwise displays 17 SNPs for a total length of 3433/3435 bp, with nine of them being found in the 1815 bp shared between the two putative reverse transcriptase encoding ORF, leading to seven amino-acid substitutions.

Genomics - plastid genome

Both plastome were also obtained. Their lengths are 133,866 bp long for strain SZCZEY2166 (GenBank: PP962255) and 133,852 bp long for strain SZCZEY2167 (GenBank: PP727123). The two plastomes came out as different isomers after assembly, hence the difference of strand of the large single copy region (LSC) that can be observed between SZCZEY2166 (Fig. 8) and SZCZEY2167 (Fig. 9). There were slight differences of lengths for LSC and SSC.

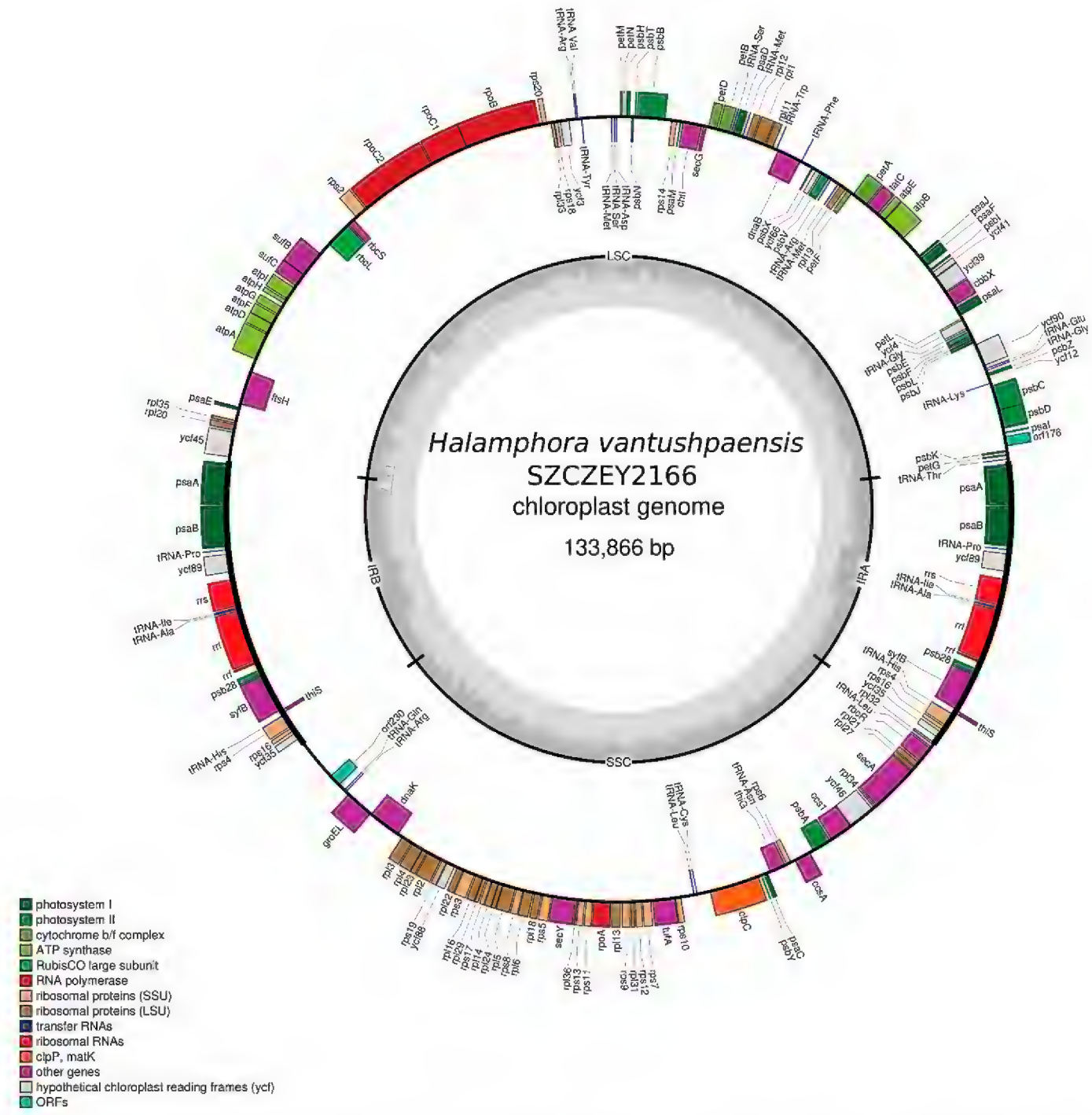

Figure 8. Map of the plastid genome of *Halamphora vantushpaensis* sp. nov. SZCZEY2166

The LSC is 61,711/61,691 bp long and display 26 indels and six SNPs. Out of these six SNPs, five were found in PCGs (*psbC*, *ycf90*, *petB*, *rps20* and *rpoC2*) and were silent only in *petB* and *rpoC2*. The short single copy (SSC) is 39,615 bp long for both strains, with no indels and 17 SNPs, all located in intergenic area. The inverted repeat had identical lengths and displayed six consecutive polymorphisms in the intergene between *rpl32* and *ycf35*. The LSC contains 70 PCGs, a single non-conserved open reading frame (ORF), and 17 tRNAs. The SSC encodes for 46 PCGs, also a single non-conserved ORF and six tRNAs. The inverted repeat IR contains 10 PCGs, three rRNA and four tRNA.

Three plastid genomes are available in GenBank for the genus *Halamphora*, all originating from the same study (Hamsher et al. 2019). In Table 3, the total lengths of these genomes and the lengths of their different compartments are compared.

Halamphora vantushpaensis has shorter LSC and SSC but its IR is consequently longer when compared to other species. The gene content of the IR is compared for all these species in Table 4. The restricted set of conserved genes found among *H. calidilacuna* or *H. americana* and which consists of a single PCG (*ycf89*), three tRNA and three rRNA seems to be shared by many unrelated species and genera such as *Navicula veneta* Kützing 1844 or *Tryblionella apiculata* Gregory 1857 (Gastineau et al. 2021a). As with *H. americana*, an extension of the IR may result from the presence of non-conserved ORF or putative genes of plasmid origin, as exemplified by its ORF9 and the putative

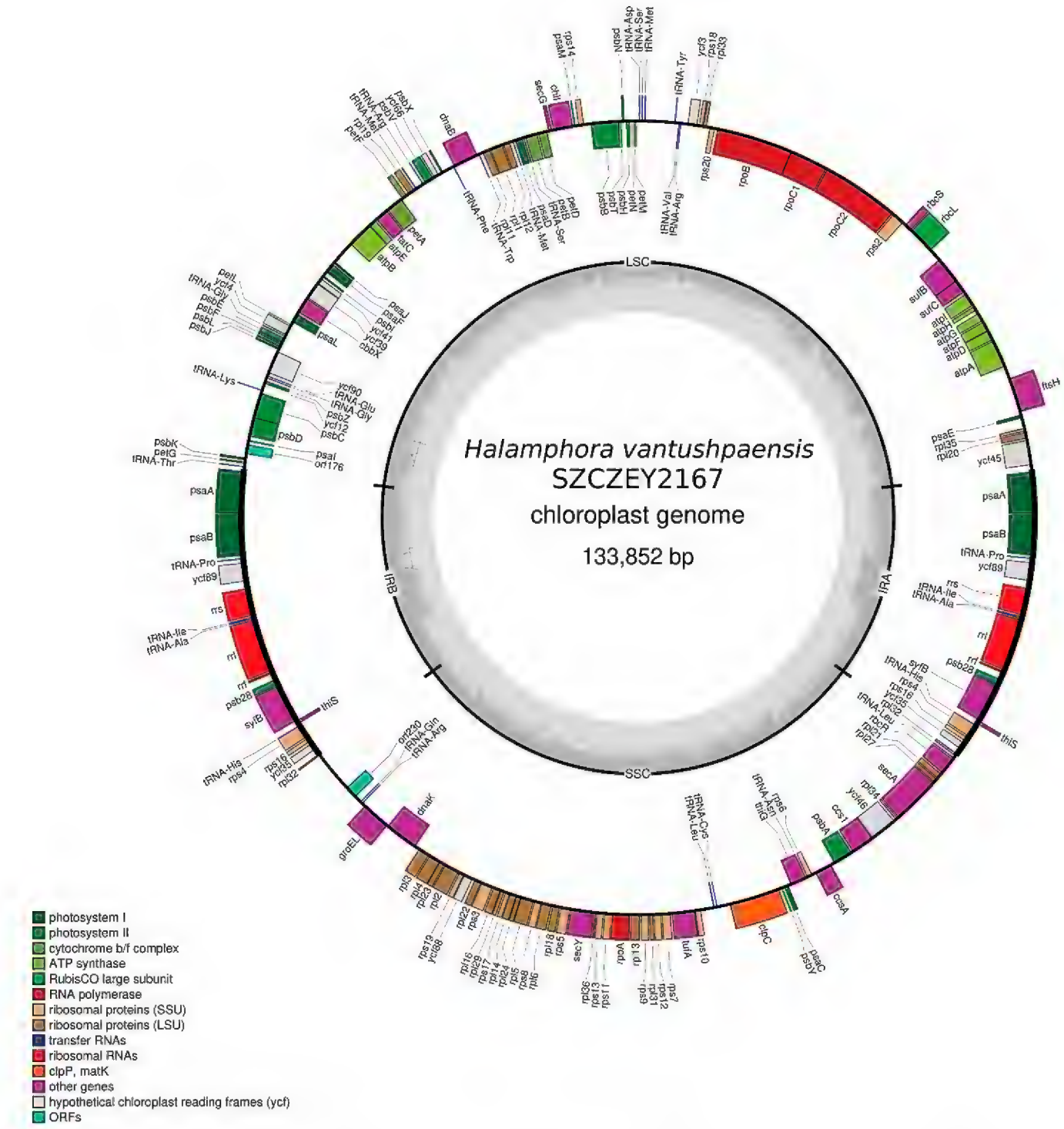

Figure 9. Map of the plastid genome of *Halamphora vantushpaensis* sp. nov. SZCZEY2167

Table 3. Lengths (in bp) of the different compartments of the plastid genomes of four species of *Halamphora* spp.

Species	Length of the LSC	Length of the SSC	Length of the IR	Total length
Halamphora calidilacuna	82,227	49,698	9,407	150,739
Halamphora americana	77,289	44,724	10,269	142,551
Halamphora coffeiformis	64,938	41,485	7,752	121,927
Halamphora vantushpaensis	61,705/61,691	39,615/39,615	16,273	133,866/133,852

Table 4. Gene composition of the inverted repeats of the plastid genomes of four species of *Halamphora* spp. Genes highlighted in bold for *Halamphora vantushpaensis* sp. nov. are genes found in the LSC in other species. Genes in bold italic concern genes usually found in the SSC. Genes marked by an asterisk are non-conserved genes of probable plasmidic origin.

Species	Gene composition of the IR
Halamphora calidilacuna	*tRNA-Pro, ycf89, rrs, tRNA-Ile, tRNA-Ala, rrl, rrf*
Halamphora americana	*tRNA-Pro, ycf89, ORF9*, rrs, tRNA-Ile, tRNA-Ala, rrl, rrf, tyrC**
Halamphora coffeiformis	*tRNA-Pro, ycf89, rrs, tRNA-Ile, tRNA-Ala, rrl, rrf, ycf35*
Halamphora vantushpaensis	**psaA, psaB**, *tRNA-Pro, ycf89, rrs, tRNA-Ile, tRNA-Ala, rrl, rrf,* ***psb28, syfb, thiS, tRNA-His, rps4, rps16, ycf35, rpl32***

integrase/recombinase encoded by the gene labelled as *tyrC* by Hamsher et al. (2019). The case of *H. vantushpaensis* is entirely different in the sense that the extension of the IR is a consequence of the incorporation of several conserved PCGs plus one tRNA. When compared with the gene content of the other species, it appears that this extension has been done at the expense of both the LSC and the SSC, which distinguishes it from species like *Climaconeis* spp. (Gastineau et al. 2021b) among which the IR seemed to have only taken over the SSC. Indeed, among the other *Halamphora* spp., *psaA* and *psaB* are located in the LSC while the other genes belong to the SSC in which they form a well-conserved cluster.

Molecular maximum likelihood phylogeny

Fig. 10 presents the *Halamphora* clade as a sub-tree derived from the three-gene inferred phylogeny. The complete three-gene tree and the *rbcL*-only tree can be found as indicated in the data availability statement. In the three-gene tree, *H. vantushpaensis* strains appear as a highly supported (99%) long-branched sister group to a larger cluster composed of 18 *Halamphora* species, namely *H. subacutiuscula*, *H. angustiformis*, *H. foramina*, *H. sydowii*, *H. tumida*, *H. witkowskii*, *H. bonnewillensis*, *H. americana*, *H. calidilacuna*, *H. intramaritima*, *H. incelebrata*, *H. banzuensis*, *H. bistriata*, *H. pertusa*, *H. subtropica* plus three unidentified *Halamphora* species. The topology of the *rbcL*-inferred tree slightly differs regarding the species sister to *H. vantushpaensis*, which are, in this case, *H. angustiformis* (bv = 93) and *H. subacutiuscula* (bv = 96). These strains are further nested in a clade with *H. maritima*, *H. pecensa*, *"Amphora" caribeana*, *H. exilis*, *H. subtropica*, *H. pertusa*, *H. banzuensis and H. bistriata* with low support (bv < 50) and, together with these, sister to *H. tumida*, *H. witkowskii*, *H. bonnewillensis*, *H. americana*, *H. calidilacuna*, *H. intramaritima*, *H. incelebrata*, *H. foramina*, *H. sydowii* and *Halamphora* sp. SZCZCH45

Discussion

Morphological comparison with similar taxa

Halamphora vantushpaensis sp. nov. is a new species, characterized through the extensive study of two distinct cultivated clones as well as examination of wild samples. The findings indicate that the morphological characteristics of *H. vantushpaensis* can strongly vary and that LM observations might not be sufficient. *Halamphora atacamana* (Patrick) Levkov, *H. borealis* (Kützing) Levkov, *H. gasseae* Levkov, and *H. salinicola* Levkov and Diaz have been identified as the most similar species. In terms of outline, *H. borealis* exhibits a semi-lanceolate shape similar to *H. vantushpaensis* (Table 1). However, distinguishing features of *H. vantushpaensis* such as the larger ventral side and the indistinct striae with a higher density (more than 27 striae in 10 µm) set it apart. Additionally, SEM images reveal differences in striae composition between *H. vantushpaensis* and *H. borealis*, further supporting their taxonomic differentiation. *Halamphora atacamana* exhibits a similar outline, especially to smaller specimens of *H. vantushpaensis*, with slightly protracted valve endings; however, larger specimens of *H. vantushpaensis* have elongated valve ends. *Halamphora atacamana* also

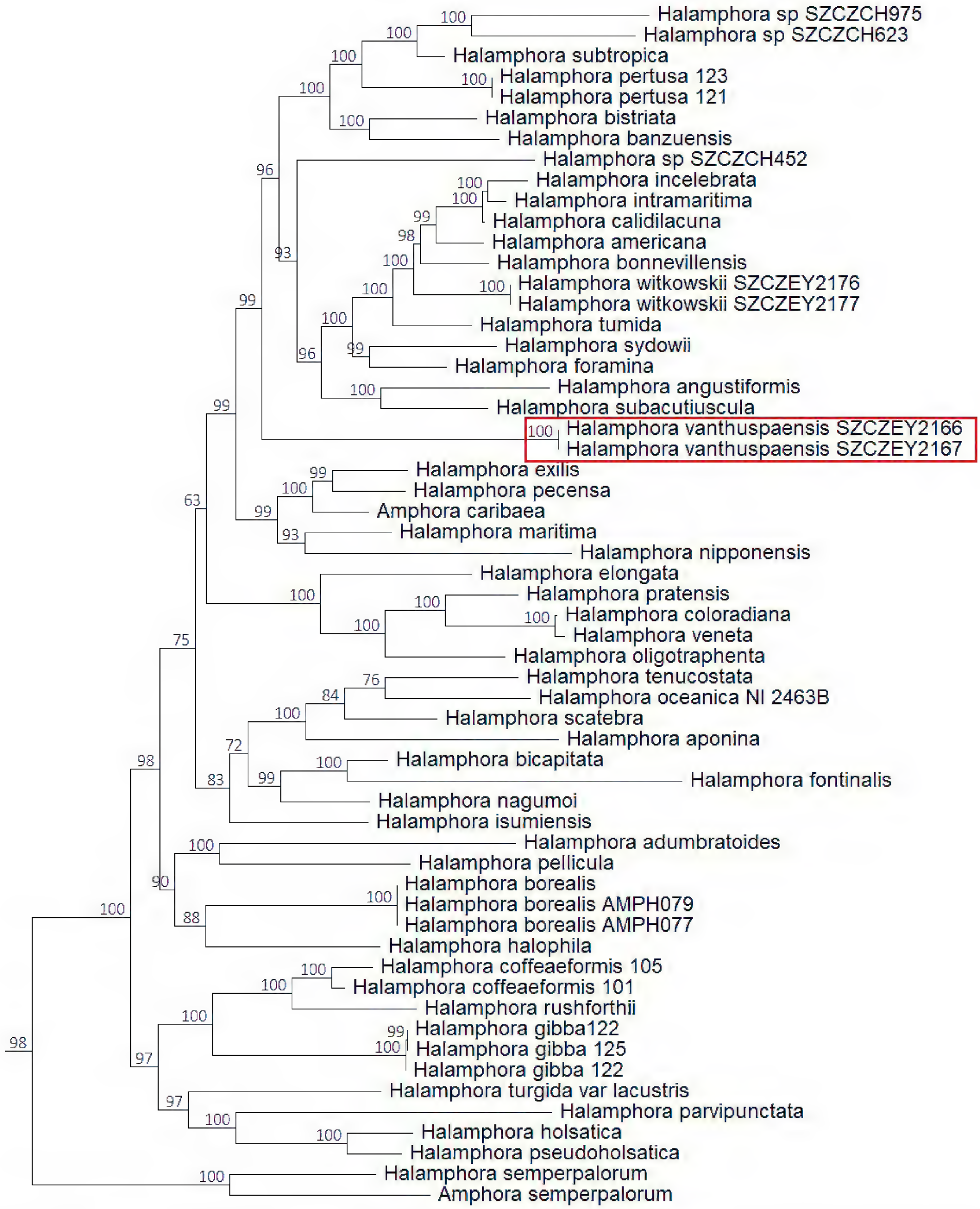

Figure 10. Maximum Likelihood phylogenetic tree inferred from concatenated alignments of *psbC*, *rbcL* and *18S*. The figure represents the sub-tree that contains the *Halamphora* clade.

tends to have lower stria densities (< 28 in 10 µm). *Halamphora gassae* and *H. salinicola* are further similar taxa, but both have smaller valves (< 35 µm in length, < 4.5 µm in width), lower stria densities (< 27 in 10 µm), and smoothly arched dorsal margins. Moreover, we observed that *H. salinicola* has larger

areolae on the dorsal side, one elongate areolae on the ventral side, and a raphe ledge that extends continuously over the entire length of the valve. Among other species that might be compared with *H. vantushpaensis*, *H. sardiniensis* has smaller valves and a higher stria density (> 36 striae in 10 µm) and strongly arched dorsal margin. Additionally, *H. sardiniensis* has elongate areolae on the dorsal side, one elongate areola on the ventral side, and a large central area on the ventral side (visible in SEM: Levkov 2009, pl.245, fig. 4). *Halamphora thermalis* is similar to smaller specimens of *H. vantushpaensis*. However, *H. thermalis* has a smoothly arched dorsal margin and a more visibly tumid ventral margin. In SEM, *H. thermalis* has larger irregularly rounded elongate areolae on dorsal side and rounded areolae on ventral side and areolae become smaller toward the central area ventrally. Also, the proximal raphe endings open into larger depressions (Levkov 2009, pl. 230, figs 1–6).

Genomic results and phylogenies

Initially, when comparing SZCZEY2166 and SZCZEY2167 by the means of LM, it was unclear whether or not they belonged to identical or different species, particularly because of the differences of shape of their apices. Of course, SEM brought supplementary clues of their identity, but in the end, molecular methods provided the decisive argument. With regards to this, it should be noted that within a single round of short-reads sequencing, complete nuclear rRNA clusters, mitochondrial genomes and plastid genomes were obtained, which allowed to derive accurate phylogenies. The *rbcL*-inferred phylogeny strictly positions *H. vantushpaensis* within a clade of species previously described as 'K clade' (Stepanek and Kociolek 2019). This is a noteworthy result, because this clade is known to contain species with extremely different preferences for what regards salinity, ranging from freshwater to hypersaline. As such, this clade is regarded as an interesting model to study transition between habitats, although it is noteworthy that such a transition seemed to have occurred repeatedly and independently among the genus Halamphora. The genomic approach we employed, sometimes described as 'genome-skimming', has reliably provided results on diatoms (for recent examples, see Gastineau et al. 2021c; Solak et al. 2021 ; Dąbek et al. 2022 ; Yilmaz et al. 2024). Aside from their interest in molecular phylogeny, the availability of full-length RNA operon reference sequences from duly identified organisms could become increasingly valuable with the development of long-read metabarcoding (e.g. Jamy et al. 2020). In the current case, it was interesting to see that the only polymorphism between both strains of *H. vantushpaensis* was located in one of the internal transcribed spacers, a portion that does not participate to the final 3D structure of the ribosome and as such, is more likely to display variations.

When comparing the plastomes, the low number of polymorphisms slightly misled us at first into thinking that these SNPs might only have concerned non-coding parts. Surprisingly, it was not the case, and although the number of SNPs is rather low, interestingly, several among them were not silent. However, this variability between the two strains is consequently lower than what was observed with the previously published Lake Van-species *N. vanseea* (Yılmaz et al. 2024). A gene such as *psbC* displayed three times more polymorphisms between the two strains studied at that time when compared to what was unveiled

between both strains of *H. vantushpaensis*. At the time *N. vanseea* was being investigated, it was possible to sequence the mitogenome for only one of the strains, for reasons that remain unknown but might be related to the amount of bacterial contamination in the DNA pool. This time, with *H. vantushpaensis*, sequencing of both strains was successful. We note that the *cox1* gene, which seems to be a sensitive marker for the study of diatoms at the subspecific level (Gastineau et al. 2013, 2021c; Dąbek et al. 2022) was entirely conserved in its exonic parts. The position of the *cox1* intron was perfectly conserved, unlike previous reports on other species (Gastineau et al. 2021c), but there were noticeable differences in its sequence. This is exemplified by the changes in length and sequence of the putative reverse-transcriptase it contains. Among land plants, introns in organellar genomes have been documented to be relevant as population markers (e.g. Spaniolas et al. 2010; Grosser et al. 2023), but no such work seems to exist for diatoms.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This work was co-financed by the Minister of Science under the "Regional Excellence Initiative" Program for 2024–2027 (RID/SP/0045/2024/01). Claude Lemieux and Monique Turmel were supported by grant RGPIN-2017-04506 from the Natural Sciences and Engineering Research Council of Canada (NSERC). We also acknowledge support from the CERCA Programme/Generalitat de Catalunya. The Royal Botanic Garden Edinburgh is supported by the Scottish Government's Rural and Environment Science and Analytical Services Division.

Author contributions

Conceptualisation: EY. Data curation: EY. Funding acquisition: AW. Investigation: EY, RG, CNS, EG, CL, MT, CO. Methodology: AW, RT, DGM. Project administration: CNS, AW. Supervision: AW, DGM, RT, CNS, RG. Visualisation: EY, CNS, EG. Writing - original draft: EY. Writing - review and editing: DGM, RT, CL, MT, CO, CNS, RG.

Author ORCIDs

Elif Yılmaz https://orcid.org/0000-0001-7939-1814
Romain Gastineau https://orcid.org/0000-0001-8661-5118
Cüneyt Nadir Solak https://orcid.org/0000-0003-2334-4271
Ewa Górecka https://orcid.org/0000-0003-0590-7480
Rosa Trobajo https://orcid.org/0000-0001-9498-3797
Monique Turmel https://orcid.org/0000-0001-7060-035X
Claude Lemieux https://orcid.org/0000-0001-9580-8042
Christian Otis https://orcid.org/0000-0001-9680-5863
Andrzej Witkowski https://orcid.org/0000-0003-1714-218X
David G. Mann https://orcid.org/0000-0003-0522-6802

Data availability

All sequences have been deposited and are already available on GenBank with accession numbers PP726705, PP726703, PP962256, PP962257, PP962255, PP727123. The complete 3-gene tree, the rbcL-inferred trees and the lists of sequences used for phylogeny can be found on Zenodo following this link: https://doi.org/10.5281/zenodo.12963401.

References

Akku M, Sar M, Ekmekçi F, Yoğurtçuoğlu B (2021) The discovery of a microbialite-associated freshwater fish in the world's largest saline soda lake, Lake Van (Turkey). Zoosystematics and Evolution 97: 181–189. https://doi.org/10.3897/zse.97.62120

Aydar E, Gourgaud A, Ulusoy I, Digonnet F, Labazuy P, Şen E, Bayhan HG, Kurttas T, Tolluoğlu AÜ (2003) Morphological analysis of active Mount Nemrut stratovolcano, eastern Turkey: evidences and possible impact areas of future eruption. Journal of Volcanology and Geothermal Research 123: 301–312. https://doi.org/10.1016/S0377-0273(03)00002-7

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455–477. https://doi.org/10.1089/cmb.2012.0021

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxford, England) 25(15): 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cleve PT (1895) Synopsis of the naviculoid diatoms, part II. Kongliga Svenska-Vetenskaps Akademiens Handlingar 27: 1–219.

Dąbek P, Gastineau R, Bornman TG, Lemieux C, Turmel M, Hallegraeff G, Mouget J-L and Witkowski A (2022) The blue diatom *Haslea ostrearia* from the Indian Ocean coast of South Africa, with comparative analysis of *Haslea* organellar genomes. Frontiers in Marine Science 9: 950716. https://doi.org/10.3389/fmars.2022.950716

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2020) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Molecular Biology and Evolution 37(1): 291–294. https://doi.org/10.1093/molbev/msz189

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13–15. https://doi.org/10.2307/2419362

Gastineau R, Leignel V, Jacquette B, Hardivillier Y, Wulff A, Gaudin P, Bendahmane D, Davidovich NA, Kaczmarska I, Mouget JL (2013) Inheritance of mitochondrial DNA in the Pennate diatom Haslea ostrearia (Naviculaceae) during auxosporulation suggests a uniparental transmission. Protist 164(3): 340–351. https://doi.org/10.1016/j.protis.2013.01.001

Gastineau R, Hamedi C, Baba Hamed MB, Abi-Ayad S-ME-A, Bąk M, Lemieux C, Turmel M, Dobosz S, Wróbel RJ, Kierzek A, Lange-Bertalot H, Witkowski A (2021a) Morphological and molecular identification reveals that waters from an isolated oasis in Tamanrasset (extreme South of Algerian Sahara) are colonized by opportunistic and pollution-tolerant diatom species. Ecological Indicators 121(6): 107104. https://doi.org/10.1016/j.ecolind.2020.107104

Gastineau R, Davidovich NA, Davidovich OI, Lemieux C, Turmel M, Wróbel RJ, Witkowski A (2021b) Extreme Enlargement of the Inverted Repeat Region in the Plastid Genomes of Diatoms from the Genus *Climaconeis*. International journal of molecular sciences 22(13): 7155. https://doi.org/10.3390/ijms22137155

Gastineau R, Hansen G, Poulin M, Lemieux C, Turmel M, Bardeau JF, Leignel V, Hardivillier Y, Morançais M, Fleurence J, Gaudin P, Méléder V, Cox EJ, Davidovich NA, Davidovich OI, Witkowski A, Kaczmarska I, Ehrman JM, Soler Onís E, Quintana AM, Mucko M, Mordret S, Sarno D, Jacquette B, Falaise C, Séveno J, Lindquist NL, Kemp PS, Jr.; Eker-Develi E, Konucu M, Mouget JL (2021c) *Haslea silbo*, A Novel Cosmopolitan Species of Blue Diatoms. Biology 10: 328. https://doi.org/10.3390/biology10040328

Gastineau R, Harðardóttir S, Guilmette C, Lemieux C, Turmel M, Otis C, Boyle B, Levesque RC, Gauthier J, Potvin M, Lovejoy C (2023) Mitochondrial genome sequence of the protist *Ancyromonas sigmoides* Kent, 1881 (Ancyromonadida) from the Sugluk Inlet, Hudson Strait, Nunavik, Québec. Frontiers in Microbiology 14: 1275665. https://doi.org/10.3389/fmicb.2023.1275665

Glombitza C, Stockhecke M, Schubert CJ, Vetter A, Kallmeyer J (2013) Sulfate reduction controlled by organic matter availability in deep sediment cores from the saline, alkaline Lake Van (Eastern Anatolia, Turkey). Frontiers in Microbiology 4: 209. https://doi.org/10.3389/fmicb.2013.00209

Gordon D, Green P (2013) Consed: A graphical editor for next-generation sequencing. Bioinformatics 29(22): 2936–2937. https://doi.org/10.1093/bioinformatics/btt515

Grosser MR, Sites SK, Murata MM, Lopez Y, Chamusco KC, Love Harriage K, Grosser JW, Graham JH, Gmitter Jr FG, Chase CD (2023). Plant mitochondrial introns as genetic markers - conservation and variation. Frontiers in plant science 14: 1116851. https://doi.org/10.3389/fpls.2023.1116851

Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. In: Smith WL, Chanley MH (Eds) Culture of Marine Invertebrate Animals. Plenum Press, New York, 26–60. https://doi.org/10.1007/978-1-4615-8714-9_3

Guiry MD, Guiry GM 30 May 2024. *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org [searched on 02 April 2024]

Hamsher SE, Keepers KG, Pogoda CS, Stepanek JG, Kane NC, Kociolek JP (2019) Extensive chloroplast genome rearrangement amongst three closely related *Halamphora* spp. (Bacillariophyceae), and evidence for rapid evolution as compared to land plants. PloS ONE 14(7): e0217824. https://doi.org/10.1371/journal.pone.0217824

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Molecular Ecology Resources 20(2): 429–443. https://doi.org/10.1111/1755-0998.13117

Kaden H, Peeters F, Lorke A, Kipfer R, Tomonaga Y, Karabiyikoglu M (2010) Impact of lake level change on deep-water renewal and oxic conditions in deep saline Lake Van, Turkey. Water Resources Research 46(11): W11508. https://doi.org/10.1029/2009WR008555

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30(4): 772–780. https://doi.org/10.1093/molbev/mst010

Kempe S, Kazmierczak J, Landmann G, Konuk T, Reimer A, Lipp A (1991) Largest known microbialites discovered in Lake Van, Turkey. Nature 349: 605–608. https://doi.org/10.1038/349605a0

Kützing FT (1844) Die kieselschaligen Bacillarien oder Diatomeen. Nordhausen, 1–152. https://doi.org/10.5962/bhl.title.64360

Legler F, Krasske G (1940) Diatomeen aus dem Vansee (Armenien). Beihefte zur Botanischen Zentralblatt 60, Abteilung B, 335–345.

Levkov Z (2009) *Amphora sensu lato*. Diatoms of Europe 5: 1–916.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37(5): 1530–1534. https://doi.org/10.1093/molbev/msaa015

North SM, Stockhecke M, Tomonaga Y Mackay AW (2018) Analysis of a fragmentary diatom record from Lake Van (Turkey) reveals substantial lake-level variability during previous interglacials MIS7 and MIS5e. Journal of Paleolimnology 59: 119–133. https://doi.org/10.1007/s10933-017-9973-z

Pogoda CS, Keepers KG, Hamsher SE, Stepanek JG, Kane NC, Kociolek JP (2019) Comparative analysis of the mitochondrial genomes of six newly sequenced diatoms reveals group II introns in the barcoding region of cox1. Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis, 30(1): 43–51. https://doi.org/10.1080/24701394.2018.1450397

Smith SA, Dunn CW (2008) Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics (Oxford, England) 24(5): 715–716. https://doi.org/10.1093/bioinformatics/btm619

Solak CN, Gastineau R, Lemieux C, Turmel M, Gorecka E, Trobajo R, Rybak M, Yılmaz E, Witkowski A (2021) *Nitzschia anatoliensis* sp. nov., a cryptic diatom species from the highly alkaline Van Lake (Turkey). PeerJ 9: e12220. https://doi.org/10.7717/peerj.12220

Spaniolas S, Bazakos C, Spano T, Zoghby C, Kalaitzis P (2010) The potential of plastid trnL (UAA) intron polymorphisms for the identification of the botanical origin of plant oils. Food Chemistry 122(3): 850–856. https://doi.org/10.1016/j.foodchem.2010.02.039

Stepanek JG, Kociolek JP (2013) Several new species of *Amphora* and *Halamphora* from the western USA. Diatom Research 28(1): 61–76. https://doi.org/10.1080/0269249X.2012.735205

Stepanek JG, Kociolek JP (2014) Molecular phylogeny of Amphora sensu lato (Bacillariophyta): an investigation into the monophyly and classification of the amphoroid diatoms. Protist 165(2): 177–195. https://doi.org/10.1016/j.protis.2014.02.002

Stepanek JG, Kociolek JP (2019) Molecular phylogeny of the diatom genera *Amphora* and *Halamphora* (Bacillariophyta) with a focus on morphological and ecological evolution. Journal of phycology 55(2): 442–456. https://doi.org/10.1111/jpy.12836

Stockhecke M, Kwiecien O, Vigliotti L, Anselmetti FS, Beer J, Çağatay MN, Channell JE, Kipfer R, Lachner J, Litt T, Pickarski N, Sturm M (2014) Chronostratigraphy of the 600,000 year old continental record of Lake Van (Turkey). Quaternary Science Reviews 104: 8–17. https://doi.org/10.1016/j.quascirev.2014.04.008

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biology and Evolution 38(7): 3022–3027. https://doi.org/10.1093/molbev/msab120

Van Heurck HF (1880) Synopsis des Diatomées de Belgique. Ducaju et Cie., 1–696. https://doi.org/10.5962/bhl.title.1990

Yilmaz E, Gastineau R, Górecka E, Solak CN, Trobajo R, Peszek Ł, Mann DG (in press) *Halamphora witkowskii* sp. nov. (Catenulaceae, Bacillariophyta), a new diatom species from the alkaline waters of Lake Van, Republic of Türkiye. Nova Hedwigia.

Yilmaz E, Mann DG, Gastineau R, Trobajo R, Solak CN, Górecka E, Turmel M, Lemieux C, Ertorun N, Witkowski A (2024) Description of *Navicula vanseea* sp. nov. (Naviculales, Naviculaceae), a new species of diatom from the highly alkaline Lake Van (Republic of Türkiye) with complete characterisation of its organellar genomes and multigene phylogeny. PhytoKeys 241: 27–48. https://doi.org/10.3897/phytokeys.241.118903